

A modified Prkopa's approach in optimum allocation in multivariate stratified random sampling

José A. Díaz-García

Department of Statistics and Computation
Universidad Autónoma Agraria Antonio Narro
25350 Buenavista, Saltillo, Coahuila, México.
jadiaz@uaaan.mx

and

Rogelio Ramos-Quiroga

Centro de Investigación en Matemáticas
Department of Probability and Statistics
Callejón de Jalisco s/n.
36240 Guanajuato, México
rramosq@ciimat.mx

Abstract

A modified Prékopa's approach is considered for the problem of optimum allocation in multivariate stratified random sampling. An example is solved by applying the proposed methodology.

Key Words: Multivariate stratified random sampling, stochastic programming, optimum allocation, integer programming, chance constraints.

Mathematics Subject Classification: 62D05, 90C15, 90C29, 90C10

1 Introduction

One of the statistical tools most commonly used in many fields of scientific research is the theory of probabilistic sampling. In diverse practical situations, the probabilistic model of stratified random sampling is frequently applied. Although there are different ways to allocate the sample in strata, the optimum allocation has been found to be a useful approach, see (Stuart, 1954), Cochran (1977), Sukhatme *et al.* (1984) and Thompson (1997).

From a multivariate point of view, there are, basically, two approaches for solving the problem of optimum allocation in multivariate stratified random sampling. When a cost function is defined as the objective function subject to certain functions of variances to be within a given region, the problem of the optimum allocation in multivariate stratified random sampling is stated as a deterministic uniojective mathematical programming problem, see Arthanari and Dodge (1981) among others. Alternatively, when the objective function is defined as some functions of variances subject to cost restrictions, the problem has been proposed implicitly and explicitly as a deterministic multiobjective mathematical programming problem, see Cochran (1977), Sukhatme *et al.* (1984) and Díaz-García and Ulloa (2008).

On the other hand, Prékopa (1978) considers the approach wherein population variances are random variables and formulated the corresponding optimum allocation problem as a stochastic (or probabilistic) mathematical programming problem, termed specifically chance constraints approach, see Charnes and Cooper (1963). Namely, Prékopa (1978) minimizes

a cost function subject to inequality restrictions in terms of the estimated variance of the stratified mean of each characteristic, restrictions that are allowed to be violated with certain probability. An alternative approach suggested by Díaz-García and Ulloa (2008) is developed by Kozak and Wang (2010) from a stochastic point of view.

This work states the optimum allocation in multivariate stratified random sampling as a stochastic integer programming problem, specifically, a modified Prékopa's approach is proposed. Section 2 includes some notation and definitions on multivariate stratified random sampling and summarizes properties on the asymptotic normality of the sample covariance matrices. The optimum allocation in multivariate stratified random sampling via chance constraints methodology is studied in Section 3. Finally an application of the approach is presented in Section 4.

2 Preliminaries on multivariate stratified random sampling

Consider a population of size N , divided into H sub-populations (strata). We wish to find a representative sample of size n and an optimum allocation rule for the strata, meeting the following requirements: i) to minimize the variance of the estimated mean, subject to a budgetary constraint; or ii) to minimize the cost subject to a constraint on the variances; this is the classical problem in optimum allocation in univariate stratified sampling, see Cochran (1977), Sukhatme *et al.* (1984) and Thompson (1997). However, if more than one characteristic (variable) is being considered, then the problem is known as optimum allocation in multivariate stratified sampling. For a formal expression of the problem of optimum allocation in multivariate stratified sampling, consider the following notation.

2.1 Notation

The subindex $h = 1, 2, \dots, H$ denotes the stratum, $i = 1, 2, \dots, N_h$ or n_h the unit within stratum h and $j = 1, 2, \dots, G$ denotes the characteristic (variable). Moreover:

N_h	Total number of units within stratum h
n_h	Number of units from the sample in stratum h
$\mathbf{Y}_h = \begin{pmatrix} \mathbf{Y}_h^1 \dots \mathbf{Y}_h^G \\ \mathbf{Y}_{h1} \dots \mathbf{Y}_{hN_h} \end{pmatrix}'$	$N_h \times G$ matrix population in stratum h ; \mathbf{Y}_{hi} is the i -th G -dimensional value of the i -th unit in stratum h
$\mathbf{y}_h = \begin{pmatrix} \mathbf{y}_h^1 \dots \mathbf{y}_h^G \\ \mathbf{y}_{h1} \dots \mathbf{y}_{hn_h} \end{pmatrix}'$	$n_h \times G$ matrix sample in stratum h ; \mathbf{y}_{hi} is the i -th element of the G -dimensional random sample in stratum h
y_{hi}^j	Value obtained for the i -th unit in stratum h of the j -th characteristic
$\mathbf{n} = (n_1, \dots, n_H)'$	Vector of the number of units in the sample
$W_h = \frac{N_h}{N}$	Relative size of stratum h
$\bar{Y}_h^j = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}^j$	Population mean in stratum h of the j -th characteristic
$\bar{\mathbf{Y}}_h = (\bar{Y}_h^1, \dots, \bar{Y}_h^G)'$	Population mean vector in stratum h

$\bar{y}_h^j = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}^j$	Sample mean in stratum h of the j -th characteristic
$\bar{\mathbf{y}}_h = (\bar{y}_h^1, \dots, \bar{y}_h^G)'$	Sample mean vector in stratum h
$\bar{\mathbf{y}}_{ST}^j = \sum_{h=1}^H W_h \bar{y}_h^j$	Estimator of the population mean in multivariate stratified sampling for the j -th characteristic
$\bar{\mathbf{y}}_{ST} = (\bar{y}_{ST}^1, \dots, \bar{y}_{ST}^G)'$	Estimator of the population mean vector in multivariate stratified sampling
\mathbf{S}_h	Variance-covariance matrix in stratum h
	$\mathbf{S}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)'$
	where $S_{h_{jk}}$ is the covariance in stratum h of the j -th and k -th characteristics; furthermore
	$S_{h_{jk}} = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{hi}^j - \bar{y}_h^j)(y_{hi}^k - \bar{y}_h^k)$, and
	$S_{h_{jj}} \equiv S_{hj}^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (y_{hi}^j - \bar{y}_h^j)^2$
\mathbf{s}_h	Estimator of the covariance matrix in stratum h ;
	$\mathbf{s}_h = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)'$
	where $s_{h_{jk}}$ is the sample covariance in stratum h of the j -th and k -th characteristics; furthermore
	$s_{h_{jk}} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi}^j - \bar{y}_h^j)(y_{hi}^k - \bar{y}_h^k)$, and
	$s_{h_{jj}} \equiv s_{hj}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi}^j - \bar{y}_h^j)^2$
$\text{Cov}(\bar{\mathbf{y}}_{ST})$	Variance-covariance matrix of $\bar{\mathbf{y}}_{ST}$
$\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$	Estimator of the covariance matrix of $\bar{\mathbf{y}}_{ST}$, it is denoted as $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \equiv \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$, and defined as
	$= \begin{pmatrix} \widehat{\text{Var}}(\bar{y}_{ST}^1) & \widehat{\text{Cov}}(\bar{y}_{ST}^1, \bar{y}_{ST}^2) & \cdots & \widehat{\text{Cov}}(\bar{y}_{ST}^1, \bar{y}_{ST}^G) \\ \widehat{\text{Cov}}(\bar{y}_{ST}^2, \bar{y}_{ST}^1) & \widehat{\text{Var}}(\bar{y}_{ST}^2) & \cdots & \widehat{\text{Cov}}(\bar{y}_{ST}^2, \bar{y}_{ST}^G) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\text{Cov}}(\bar{y}_{ST}^G, \bar{y}_{ST}^1) & \widehat{\text{Cov}}(\bar{y}_{ST}^G, \bar{y}_{ST}^2) & \cdots & \widehat{\text{Var}}(\bar{y}_{ST}^G) \end{pmatrix}$
	$= \sum_{h=1}^H \frac{W_h^2 \mathbf{s}_h}{n_h} - \sum_{h=1}^H \frac{W_h \mathbf{s}_h}{N}$
$\widehat{\text{Cov}}(\bar{y}_{ST}^j, \bar{y}_{ST}^k)$	Estimated covariance of \bar{y}_{ST}^j and \bar{y}_{ST}^k where
	$\widehat{\text{Cov}}(\bar{y}_{ST}^j, \bar{y}_{ST}^k) \equiv \widehat{\text{Cov}}(\bar{y}_{ST}^j, \bar{y}_{ST}^k)$, with
	$\widehat{\text{Cov}}(\bar{y}_{ST}^j, \bar{y}_{ST}^k) = \sum_{h=1}^H \frac{W_h^2 s_{h_{jk}}}{n_h} - \sum_{h=1}^H \frac{W_h s_{h_{jk}}}{N}$, and
	$\widehat{\text{Cov}}(\bar{y}_{ST}^j, \bar{y}_{ST}^j) \equiv \widehat{\text{Var}}(\bar{y}_{ST}^j) = \sum_{h=1}^H \frac{W_h^2 s_{hj}^2}{n_h} - \sum_{h=1}^H \frac{W_h s_{hj}^2}{N}$
c_h	Cost per G -dimensional sampling unit in stratum h and let $\mathbf{c} = (c_1, \dots, c_G)'$.

Where if $\mathbf{a} \in \Re^G$, \mathbf{a}' denotes the transpose of \mathbf{a} .

2.2 Asymptotic normality

Now, the asymptotic distribution of the estimator, \mathbf{s}_h , of the covariance matrix is stated. First, consider the following notation and definitions.

A detailed discussion of operator “vec”, “vech”, Moore-Penrose inverse, Kronecker product, commutation matrix and duplication matrix may be found in Magnus and Neudecker (1988), among many others. For convenience, some notation is introduced, although in general it adheres to standard notation.

For all matrix \mathbf{A} there exists a unique matrix \mathbf{A}^+ which is termed *Moore-Penrose inverse* of \mathbf{A} .

Let \mathbf{A} be an $m \times n$ matrix and \mathbf{B} a $p \times q$ matrix. The $mp \times nq$ matrix defined by

$$\begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

is termed the *Kronecker product* (also termed tensor product or direct product) of \mathbf{A} and \mathbf{B} and written $\mathbf{A} \otimes \mathbf{B}$. Let \mathbf{C} be an $m \times n$ matrix and \mathbf{C}_j its j -th column, then $\text{vec } \mathbf{C}$ is the $mn \times 1$ vector

$$\text{vec } \mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_n \end{bmatrix}.$$

The vectors $\text{vec } \mathbf{C}$ and $\text{vec } \mathbf{C}'$ clearly contain the same mn components, but in different order. Therefore there exists a unique $mn \times mn$ permutation matrix which transforms $\text{vec } \mathbf{C}$ into $\text{vec } \mathbf{C}'$. This matrix is termed the *commutation matrix* and is denoted \mathbf{K}_{mn} (If $m = n$, it is often written \mathbf{K}_n instead of \mathbf{K}_{mn}). Hence

$$\mathbf{K}_{mn} \text{vec } \mathbf{C} = \text{vec } \mathbf{C}'.$$

Similarly, let \mathbf{B} be a square $n \times n$ matrix. Then $\text{vech } \mathbf{B}$ (also denoted as $\mathbf{v}(\mathbf{B})$) shall denote the $n(n+1)/2 \times 1$ vector that is obtained from $\text{vec } \mathbf{B}$ by eliminating all supradiagonal elements of \mathbf{B} . If $\mathbf{B} = \mathbf{B}'$, $\text{vech } \mathbf{B}$ contains only the distinct elements of \mathbf{B} , then there exists a unique $n^2 \times n(n+1)/2$ matrix termed *duplication matrix*, which is denoted by \mathbf{D}_n , such that $\mathbf{D}_n \text{vech } \mathbf{B} = \text{vec } \mathbf{B}$ and $\mathbf{D}_n^+ \text{vec } \mathbf{B} = \text{vech } \mathbf{B}$. Finally, note that $(\text{vech } \mathbf{B})' \equiv \text{vech}' \mathbf{B}$.

Now, with the above mathematical tools and based in the extension given in Hájek (1961), the multivariate version of Hájek's theorem is restated in terms of sampling theory terminology, which is explained in detail in Díaz García and Ramos-Quiroga (2011).

Lemma 2.1. *Let Ξ_ν be a $G \times G$ symmetric random matrix defined as*

$$\Xi_\nu = \frac{1}{n_\nu - 1} \sum_{i=1}^{n_\nu} (\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)'$$

Suppose that for $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$, any vector of constants, $k = G(G+1)/2$,

$$\boldsymbol{\lambda}' (\mathbf{M}_\nu^4 - \text{vech } \mathbf{S}_\nu \text{vech}' \mathbf{S}_\nu) \boldsymbol{\lambda} \geq \epsilon \max_{1 \leq \alpha \leq k} \left[\lambda_\alpha^2 \mathbf{e}_k^{\alpha'} (\mathbf{M}_\nu^4 - \text{vech } \mathbf{S}_\nu \text{vech}' \mathbf{S}_\nu) \mathbf{e}_k^\alpha \right], \quad (1)$$

where $\mathbf{e}_k^\alpha = (0, \dots, 0, 1, 0, \dots, 0)'$ is the α -th vector of the canonical base of \mathbb{R}^k , $\epsilon > 0$ and independent of $\nu > 1$ and

$$\mathbf{M}_\nu^4 = \frac{1}{N_\nu} \mathbf{D}_G^+ \left[\sum_{i=1}^{N_\nu} (\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)' \otimes (\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)' \right] \mathbf{D}_G^{+'},$$

is the fourth central moment. Assume that $n_\nu \rightarrow \infty$, $N_\nu - n_\nu \rightarrow \infty$, $N_\nu \rightarrow \infty$, and that, for all $j = 1, \dots, G$,

$$\left[\lim_{\nu \rightarrow \infty} \left(\frac{n_\nu}{N_\nu} \right) = 0 \right] \Rightarrow \lim_{\nu \rightarrow \infty} \frac{\max_{1 \leq i_1 < \dots < i_{n_\nu} \leq N_\nu} \sum_{\beta=1}^{n_\nu} \left[\left(y_{\nu i_\beta}^j - \bar{Y}_\nu^j \right)^2 - S_{\nu j}^2 \right]^2}{N_\nu \left[m_{\nu j}^4 - \left(S_{\nu j}^2 \right)^2 \right]} = 0, \quad (2)$$

where

$$m_{\nu j}^4 = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} \left(y_{\nu i}^j - \bar{y}_\nu^j \right)^4.$$

Then, $\text{vech } \Xi_\nu$ is asymptotically normal distributed as

$$\text{vech } \Xi_\nu \xrightarrow{d} \mathcal{N}_k(\mathbf{E}(\text{vech } \Xi_\nu), \text{Cov}(\text{vech } \Xi_\nu)),$$

with

$$\mathbf{E}(\text{vech } \Xi_\nu) = \frac{n_\nu}{n_\nu - 1} \text{vech } \mathbf{S}_\nu, \quad (3)$$

and

$$\text{Cov}(\text{vech } \Xi_\nu) = \frac{n_\nu}{(n_\nu - 1)^2} (\mathbf{M}_\nu^4 - \text{vech } \mathbf{S}_\nu \text{vech}' \mathbf{S}_\nu). \quad (4)$$

n_ν is the sample size for a simple random sample from the ν -th population of size N_ν .

Then:

Theorem 2.1. Under assumptions in Lemma 2.1, the sequence of sample covariance matrices \mathbf{s}_ν are such that $\text{vech } \mathbf{s}_\nu$ has an asymptotical normal with asymptotic mean and covariance matrix given by (3) and (4), respectively.

Proof. This follows immediately from Lemma 2.1, only observe that

$$\begin{aligned} \mathbf{s}_\nu &= \frac{1}{n_\nu - 1} \sum_{i=1}^{n_\nu} (\mathbf{y}_{\nu i} - \bar{\mathbf{y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{y}}_\nu)' \\ &= \Xi - \frac{n_\nu}{n_\nu - 1} (\bar{\mathbf{y}}_\nu - \bar{\mathbf{Y}}_\nu)(\bar{\mathbf{y}}_\nu - \bar{\mathbf{Y}}_\nu)', \end{aligned}$$

where

$$\frac{n_\nu}{n_\nu - 1} \rightarrow 1 \quad \text{and} \quad (\bar{\mathbf{y}}_\nu - \bar{\mathbf{Y}}_\nu)(\bar{\mathbf{y}}_\nu - \bar{\mathbf{Y}}_\nu)' \rightarrow 0 \quad \text{in probability.} \quad \square$$

□

Remark 2.1. Observe that it is possible to find the asymptotic distribution of $\text{vec } \mathbf{S}_\nu$, but this asymptotic normal distribution is singular, because $\text{Cov}(\text{vec } \mathbf{S}_\nu)$ is singular. This is due to the fact $\text{Cov}(\text{vec } \mathbf{S}_\nu)$ is the $G^2 \times G^2$ covariance matrix in the asymptotic distribution of $\text{vec } \mathbf{S}_\nu$ and, because \mathbf{S}_ν is symmetric, then $\text{vec } \mathbf{S}_\nu$ has repeated elements. In this case, $\text{vec } \mathbf{S}_\nu$ is asymptotically normally distributed as (see Muirhead (1982))

$$\text{vec } \mathbf{S}_\nu \xrightarrow{d} \mathcal{N}_{G^2}(\mathbf{E}(\text{vec } \mathbf{\Xi}_\nu), \text{Cov}(\text{vec } \mathbf{\Xi}_\nu)),$$

where

$$\mathbf{E}(\text{vec } \mathbf{\Xi}_\nu) = \frac{n_\nu}{n_\nu - 1} \text{vec } \mathbf{S}_\nu,$$

$$\text{Cov}(\text{vec } \mathbf{\Xi}_\nu) = \frac{n_\nu}{(n_\nu - 1)^2} (\mathfrak{M}_\nu^4 - \text{vec } \mathbf{S}_\nu \text{vec}' \mathbf{S}_\nu),$$

and

$$\mathfrak{M}_\nu^4 = \frac{1}{N_\nu} \left[\sum_{i=1}^{N_\nu} (\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)' \otimes (\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)(\mathbf{y}_{\nu i} - \bar{\mathbf{Y}}_\nu)' \right]. \quad \square$$

The following assertion is an immediate consequence of Theorem 2.1.

Theorem 2.2. Let $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$ be the estimator of the covariance matrix of $\bar{\mathbf{y}}_{ST}$, then

$$\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) = \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right) \text{vech } \mathbf{s}_h$$

is asymptotically normally distributed; furthermore

$$\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \xrightarrow{d} \mathcal{N}_k \left(\mathbf{E} \left(\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right), \text{Cov} \left(\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) \right), \quad (5)$$

where

$$\mathbf{E} \left(\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) = \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right) \frac{n_h}{n_h - 1} \text{vech } \mathbf{s}_h, \quad (6)$$

$$\begin{aligned} \text{Cov} \left(\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) &= \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right)^2 \frac{n_h}{(n_h - 1)^2} (\mathbf{M}_h^4 - \text{vech } \mathbf{s}_h \text{vech}' \mathbf{s}_h), \end{aligned} \quad (7)$$

and

$$\mathbf{M}_h^4 = \frac{1}{N_h} \mathbf{D}_G^+ \left[\sum_{i=1}^{N_h} (\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)' \otimes (\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{Y}}_h)' \right] \mathbf{D}_G^{+'}.$$

Finally, note that the asymptotically normal distributions of $\text{vech } \mathbf{s}_h$, $\text{vec } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$ and $\text{vech } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$ are in terms of the parameters $\bar{\mathbf{Y}}_h$, $\text{vech } \mathbf{s}_h$, \mathfrak{M}_h^4 and \mathbf{M}_h^4 ; then, from Rao (1973, iv), pp. 388-389), approximations of asymptotic distributions can be obtained, making the following substitutions

$$\bar{\mathbf{Y}}_h \rightarrow \bar{\mathbf{y}}_h, \quad \text{vech } \mathbf{s}_h \rightarrow \text{vech } \mathbf{s}_h, \quad \mathfrak{M}_h^4 \rightarrow \mathbf{m}_h^4 \quad \text{and} \quad \mathbf{M}_h^4 \rightarrow \mathbf{m}_h^4 \quad (8)$$

where

$$\mathbf{m}_h^4 = \frac{1}{n_h} \mathbf{D}_G^+ \left[\sum_{i=1}^{n_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \otimes (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \right] \mathbf{D}_G^{+'},$$

and

$$\mathbf{m}_h^4 = \frac{1}{n_h} \left[\sum_{i=1}^{n_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \otimes (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \right].$$

3 Modified Prékopa's approach

Optimum allocation in multivariate stratified random sampling was proposed as the following deterministic mathematical programming problem

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_0 \\ & \text{subject to} \\ & \widehat{\text{Var}}(\bar{\mathbf{y}}_{ST}^j) \leq v_0^j, \quad j = 1, 2, \dots, G \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}, \end{aligned} \tag{9}$$

where v_0^j are desired precisions assigned to the variances of the sample mean $\widehat{\text{Var}}(\bar{\mathbf{y}}_{ST}^j)$, $j = 1, 2, \dots, G$. This approach has been treated in detail by Arthanari and Dodge (1981).

From a stochastic point of view of (9), Prékopa (1978) proposes the following chance constraints mathematical program

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_0 \\ & \text{subject to} \\ & \text{P} \left(\widehat{\text{Var}}(\bar{\mathbf{y}}_{ST}^j) \leq v_0^j \right) \geq p_0, \quad j = 1, 2, \dots, G \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N}, \end{aligned} \tag{10}$$

where $0 \leq p_0 \leq 1$ is a specified probability.

The present work considers the following alternative chance constraints mathematical programming problem

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_0 \\ & \text{subject to} \\ & \text{P} \left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) < \Delta \right) \geq p_0 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N} \\ & \text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \xrightarrow{d} \mathcal{N}_k \left(\text{E} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right), \text{Cov} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) \right), \end{aligned} \tag{11}$$

where $\Delta > \mathbf{0}$ is a constant matrix.

From Díaz-García and Ulloa (2008), note that $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$ is an explicit function of \mathbf{n} , and so it must be denoted as $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \equiv \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}))$. In addition, assume that $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}))$ is a positive definite matrix for all \mathbf{n} , $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n})) > \mathbf{0}$. Now, let \mathbf{n}_1 and

\mathbf{n}_2 be two possible values of the vector \mathbf{n} and, recall that, for \mathbf{A} and \mathbf{B} positive definite matrices, $\mathbf{A} > \mathbf{B} \Leftrightarrow \mathbf{A} - \mathbf{B} > \mathbf{0}$. Hence, there exists a function f such that: $f : \mathcal{S} \rightarrow \mathbb{R}$,

$$\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}_1)) < \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}_2)) \Leftrightarrow f\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}_1))\right) < f\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n}_2))\right) \quad (12)$$

with $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}(\mathbf{n})) \in \mathcal{S} \subset \mathbb{R}^{G(G+1)/2}$ and \mathcal{S} is the set of positive definite matrices.

Then, (11) can be reduced to the following chance constraints mathematical program

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{c}'\mathbf{n} + c_0 \\ & \text{subject to} \\ & \text{P}\left(f\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right) \leq \tau\right) \geq p_0 \\ & 2 \leq n_h \leq N_h, \quad h = 1, 2, \dots, H \\ & n_h \in \mathbb{N} \\ & \text{vech}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right) \xrightarrow{d} \mathcal{N}_k\left(\text{E}\left(\text{vech}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right)\right), \text{Cov}\left(\text{vech}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right)\right)\right). \end{aligned} \quad (13)$$

There are many possibilities for the definition of $f(\cdot)$, see Díaz-García and Ulloa (2008). In particular, it is of interest when $f = \text{tr}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right)$. Among many others options, it is also interesting the case when $f = \left|\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right|$ in (13) which is described in detail in Section 4, although its application in a real problem poses some algorithmic and numerical challenges still under study.

4 Application

Lets consider the results of a forest survey conducted in Humboldt County, California, originally reported in Arvanitis and Afonja (1971). The population was subdivided into nine strata on the basis of the timber volume per unit area, as determined from aerial photographs. The two variables included in this example are the basal area (BA)¹ in square feet, and the net volume in cubic feet (Vol.), both expressed on a per acre basis. The variances, covariances and the number of units within stratum h are listed in Table 1.

For this example, the matrix optimisation problem under approach (13) is

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{n}'\mathbf{c} + c_0 \\ & \text{subject to} \\ & \text{P}\left(f\left(\begin{pmatrix} \widehat{\text{Var}}(\bar{y}_{ST}^1) & \widehat{\text{Cov}}(\bar{y}_{ST}^1, \bar{y}_{ST}^2) \\ \widehat{\text{Cov}}(\bar{y}_{ST}^2, \bar{y}_{ST}^1) & \widehat{\text{Var}}(\bar{y}_{ST}^2) \end{pmatrix}\right) \leq \tau\right) \geq p_0 \\ & \sum_{h=1}^9 n_h = 1000 \\ & 2 \leq n_h \leq N_h, \quad h = 1, \dots, 9 \\ & \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \xrightarrow{d} \mathcal{N}_{2 \times 2}\left(\text{E}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right), \text{Cov}\left(\text{vec}\left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})\right)\right)\right) \\ & n_h \in \mathbb{N}. \end{aligned} \quad (14)$$

¹In forestry terminology, 'Basal area' is the area of a plant perpendicular to the longitudinal axis of a tree at 4.5 feet above ground.

Table 1: Variances, covariances and the number of units within each stratum

Stratum	N_h	c_h^a	Variance		Covariance
			BA	Vol.	
1	11 131	2.5	1 557	554 830	28 980
2	65 857	3.0	3 575	1 430 600	61 591
3	106 936	1.5	3 163	1 997 100	72 369
4	72 872	2.5	6 095	5 587 900	166 120
5	78 260	2.0	10 470	10 603 000	293 960
6	51 401	2.0	8 406	15 828 000	357 300
7	24 050	2.5	20 115	26 643 000	663 300
8	46 113	3.0	9 718	13 603 000	346 810
9	102 985	3.5	2 478	1 061 800	39 872

^aThese are simulated costs, also c_0 is taken as 0

4.1 Solution when $f(\cdot) \equiv \text{tr}(\cdot)$

Observe that by (5), (6) and (7)

$$\text{tr Cov}(\bar{\mathbf{y}}_{ST}) \sim \mathcal{N}(\text{E}(\text{tr Cov}(\bar{\mathbf{y}}_{ST})), \text{Var}(\text{tr Cov}(\bar{\mathbf{y}}_{ST})))$$

where

$$\begin{aligned} \text{E}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})) &= \sum_{j=1}^G \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right) \frac{n_h}{n_h - 1} S_{h_j}^2, \\ \text{Var}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})) &= \sum_{j=1}^G \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right)^2 \frac{n_h}{(n_h - 1)^2} (m_{h_j}^4 - (S_{h_j}^2)^2), \end{aligned}$$

and

$$m_{h_j}^4 = \frac{1}{N_h} \left[\sum_{i=1}^{N_h} (y_{hi}^j - \bar{Y}_h^j)^4 \right].$$

Standardising the function f in equation (14), it is seen that

$$\text{P} \left[\frac{\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) - \text{E}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}{\sqrt{\text{Var}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}} \leq \frac{\tau - \text{E}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}{\sqrt{\text{Var}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}} \right] \geq p_0,$$

with

$$p_0 = \Phi \left(\frac{\tau - \text{E}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}{\sqrt{\text{Var}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}} \right),$$

where $\Phi(\cdot)$, denotes the standard normal distribution function. Let e_{p_0} be the value of the standard normal random variable such that $\Phi(e_{p_0}) = p_0$, in such way that the inequality can be established as

$$\Phi \left(\frac{\tau - \text{E}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}{\sqrt{\text{Var}(\text{tr } \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}))}} \right) \geq \Phi(e_{p_0}),$$

which holds only if

$$\frac{\tau - \mathbb{E}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST}))}{\sqrt{\widehat{\text{Var}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST}))}} \geq e_{p_0},$$

or equivalently

$$\mathbb{E}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST})) + e_{p_0} \sqrt{\widehat{\text{Var}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST}))} - \tau \leq 0. \quad (15)$$

Hence, taking into account (8), the equivalent deterministic problem to the stochastic mathematical programming (14), is given by

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{n}'\mathbf{c} + c_0 \\ & \text{subject to} \\ & \widehat{\mathbb{E}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST})) + e_{p_0} \sqrt{\widehat{\text{Var}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST}))} - \tau \leq 0 \\ & \sum_{h=1}^9 n_h = 1000 \\ & 2 \leq n_h \leq N_h, \quad h = 1, \dots, 9 \\ & n_h \in \mathbb{N}. \end{aligned}$$

where

$$\widehat{\mathbb{E}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST})) = \sum_{j=1}^G \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right) \frac{n_h}{n_h - 1} s_{hj}^2, \quad (16)$$

$$\widehat{\text{Var}}(\widehat{\text{tr Cov}}(\bar{\mathbf{y}}_{ST})) = \sum_{j=1}^G \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right)^2 \frac{n_h}{(n_h - 1)^2} (m_{hj}^4 - (s_{hj}^2)^2), \quad (17)$$

and

$$m_{hj}^4 = \frac{1}{n_h} \left[\sum_{i=1}^{n_h} (y_{hi}^j - \bar{y}_h^j)^4 \right]. \quad (18)$$

Remark 4.1. Observe that the estimators \bar{y}_h^j , s_{hj}^2 and m_{hj}^4 of \bar{Y}_h^j , S_{hj}^2 and M_{hj}^4 could initially be obtained as

- i) results from a pilot (preliminary) sample or
- ii) using the corresponding values of the estimators from another set of variables, X 's, correlated to the variables Y 's.

It is important to have this in mind in the minimisation step, because for example, the n_h that appears in expression (18), is the value of n_h (fixed) used in the pilot study. Same comment for the expression of the estimator \bar{y}_h^j and s_{hj}^2 . While the n_h 's that appear in expressions (16) and (17) are the decision variables. \square

4.2 Solution when $f(\cdot) \equiv |\cdot|$

Assume the following alternative stochastic matrix mathematical programming problem

$$\begin{aligned}
& \min_{\mathbf{n}} \mathbf{n}'\mathbf{c} + c_0 \\
& \text{subject to} \\
& \text{P} \left(f \left(\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) \leq \tau \right) \geq p_0 \\
& \sum_{h=1}^9 n_h = 1000 \\
& 2 \leq n_h \leq N_h, \quad h = 1, \dots, 9 \\
& \text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \xrightarrow{d} \mathcal{N}_{G \times G} \left(\text{vech} \mathbf{0}_{G \times G}, \text{Cov} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) \right) \\
& n_h \in \mathbb{N},
\end{aligned} \tag{19}$$

where $\widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST})$

$$= \text{vech}^{-1} \left[\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) - \text{E} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) \right]$$

and vech^{-1} is the inverse function of function vech .

Then, the restriction in (19), is

$$\text{P} \left(\left| \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right| \leq \tau \right) \geq p_0$$

which for $G = 2$ and assuming that $\widehat{\text{Cov}} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right)$ is such that

$$\widehat{\text{Cov}} \left(\text{vech} \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right) = \mathbf{B} \otimes \mathbf{B} = \mathbf{N},$$

implies that

$$\text{P} \left(\left| \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right| \leq \tau |\mathbf{N}|^{1/4} \right) \geq p_0$$

where

$$\begin{aligned}
\mathbf{N} &= \sum_{h=1}^H \left(\frac{W_h^2}{n_h} - \frac{W_h}{N} \right)^2 \frac{n_h}{(n_h - 1)^2} (\mathbf{m}_h^4 - \text{vec } \mathbf{s}_h \text{vec}' \mathbf{s}_h), \\
\mathbf{m}_h^4 &= \frac{1}{n_h} \left[\sum_{i=1}^{n_h} (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \otimes (\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)(\mathbf{y}_{hi} - \bar{\mathbf{y}}_h)' \right].
\end{aligned}$$

see Remark 4.1, and

$$p_0 = \Psi \left(\tau |\mathbf{N}|^{1/4} \right),$$

with $\Psi(\cdot)$, denotes the distribution function of $\left| \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right|$, see Delannay and Caër (2000). Let r_{p_0} be the percentile of a random variable such that $\Psi(r_{p_0}) = p_0$, in such way that the inequality can be established as

$$\Psi \left(\left| \widehat{\text{Cov}}(\bar{\mathbf{y}}_{ST}) \right| \leq \tau |\mathbf{N}|^{1/4} \right) \geq \Psi(r_{p_0}),$$

which holds only if

$$\tau |\mathbf{N}|^{1/4} \geq r_{p_0},$$

where the density of $Z = \widehat{\text{Cov}}(\bar{\mathbf{y}}_{s_T})$ is, see Delannay and Caër (2000)

$$\frac{dG(z)}{dz} = g_Z(z) = \frac{1}{\sqrt{2}} \exp(z) \left[1 - \text{erf}(\sqrt{2z}) \right], \quad z \geq 0,$$

where $\text{erf}(\cdot)$ is the usual error function defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt.$$

Thus, by (8), the equivalent deterministic problem to the stochastic mathematical programming problem (19), is given by

$$\begin{aligned} & \min_{\mathbf{n}} \mathbf{n}' \mathbf{c} + c_0 \\ & \text{subject to} \\ & \tau |\mathbf{N}|^{1/4} \geq r_{p_0} \\ & \sum_{h=1}^9 n_h = 1000 \\ & 2 \leq n_h \leq N_h, \quad h = 1, \dots, 9 \\ & n_h \in \mathbb{N}, \end{aligned}$$

Table 2 includes the optimum allocation for each characteristic, BA and Vol (the second and third rows) from a deterministic point of view. Also appear (on fourth and fifth rows) the optimal allocations via the deterministic problem (9), identified in the table with the name Prékopa, and the deterministic version of (13) when $f(\cdot) = \text{tr}(\cdot)$. These results are presented in their stochastic version in the 7-10th rows. The last three columns show the minimum values of the individual variances for the respective optimum allocations and the cost identified by each method. The results were computed using the commercial software Hyper LINGO/PC, release 6.0, see Winston (1995). The default optimisation methods used by LINGO to solve the nonlinear integer optimisation programs are Generalised Reduced Gradient (GRG) and branch-and-bound methods, see Bazaraa *et al.* (2006). Finally, note that, for this sampling study, there is not a great discrepancy between the different methods among the sizes of the strata. And for the multivariate solutions, the biggest cost difference appears in the deterministic version of Prékopa's method.

5 Conclusions

There is a vast literature on the problem of sample allocation in stratified sampling. A natural approach considers a cost minimisation problem subject to variance restrictions. This paper follows Prékopa's approach by setting the problem into the area of stochastic optimization. It is recognized that this is a more realistic approach because, in general, the population variances of the strata are unknown and therefore requires estimating them. As a result, problem (9) really falls within the scope of stochastic mathematical programming which incorporates the inherent uncertainty of estimators in a natural way.

The approach is not without its drawbacks, it is difficult to give general rules for electing the value function $f(\cdot)$, potentially there are an infinite number of possibilities. In this paper we have chosen to work with $f(\mathbf{A}) = |\mathbf{A}|$ and $f(\mathbf{A}) = \text{tr}(\mathbf{A})$ which can be interpreted as a generalised variance and as an average variance respectively. However, the responsibility for the selection or definition of that function, lies wholly with the expert in the field of application.

Table 2: Sample sizes and estimator of variances for the different allocation rules

Allocation	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	n_9	$\widehat{\text{Var}}(\overline{y}_{ST}^1)$	$\widehat{\text{Var}}(\overline{y}_{ST}^2)$	Cost
Deterministic approach												
BA ^a	10	78	171	123	194	114	75	90	94	5.599	5766.161	2225.5
Vol ^b	6	51	139	123	204	163	90	109	64	6.502	5499.996	2194.0
Prékopa ^c	10	78	171	123	194	114	75	90	94	5.599	5766.161	2225.5
$\widehat{\text{tr Cov}}(\overline{\mathbf{y}}_{ST})^d$	6	47	127	114	186	149	80	102	59	7.071	5992.921	2014.0
Stochastic approach^e												
BA	10	79	168	125	196	117	76	91	95	5.939	5693.354	2248.0
Vol	6	48	129	113	189	150	82	102	60	6.988	5933.759	2034.0
Prékopa	11	79	169	123	196	117	78	91	96	5.921	5680.571	2034.0
$\widehat{\text{tr Cov}}(\overline{\mathbf{y}}_{ST})$	6	48	129	114	188	151	81	102	60	6.988	5933.752	2034.0

^aWith $v_0^1 = 6$

^bWith $v_0^2 = 6000$

^cWith $v_0^1 = 6$ and With $v_0^2 = 6000$

^dWith $\tau = 6000$

^eWith $p_0 = 0.50$

Acknowledgments

This research work was partially supported by IDI-Spain, Grants No. FQM2006-2271 and MTM2008-05785, supported also by CONACYT Grant CB2008 Ref. 105657. This paper was written during J. A. Díaz-García's stay as a visiting professor at the Department of Probability Statistics of the Center of Mathematical Research, Guanajuato, México.

REFERENCES

- Arvanitis, L. G., and Afonja, B., 1971. Use of the generalized variance and the gradient projection method in multivariate stratified sampling. *Biometrics*, 27, 119-127.
- Arthanari, T. S., Dodge Y., 1981. *Mathematical Programming in Statistics*. John Wiley & Sons, New York.
- Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 2006. *Nonlinear Programming: Theory and Algorithms*, 3rd Edition. Wiley-Interscience.
- Charnes, A., Cooper, W. W., 1963. Deterministic equivalents for optimizing and satisficing under chance constraints. *Operation Research*, 11, 18-39.
- Cochran, W. G. 1977. *Sampling Techniques*. Wiley, New York.
- Delannay, R. and Caër, G. Le., 2000. Distribution of the determinant of a random real-symmetric matrix from the Gaussian orthogonal ensemble. *Physical Review E* 62, 1526-1536.
- Díaz García, J. A., Ramos-Quiroga, R., 2011. Optimum allocation in multivariate stratified random sampling: Stochastic matrix optimisation. <http://arxiv.org/abs/1105.3224>.
- Díaz-García, J. A., and Ulloa, C. L. 2008. Multi-objective optimisation for optimum allocation in multivariate stratified sampling. *Survey Methodology*, 34(2), 215-222.

- Giri, N. C., 1977. Multivariate Statistical Inference. Academic Press, New York.
- Hájek, J. 1961. Some extensions of the Wald-Wolfowitz-Noether theorem. The Annals of Mathematical Statistics 32, 506-523.
- Kozak, M., and Wang, H. Y. 2010. On stochastic optimization in sample allocation among strata. METRON, LXVIII (1), 95-103.
- Magnus, J. R., Neudecker, H. 1988. Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley & Sons, New York.
- Muirhead, R. J., 1982. Aspects of Multivariate Statistical Theory. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc.
- Prékopa, A., 1978. The use of stochastic programming for the solution of some problems in statistics and probability. Technical Summary report #1834. University of Wisconsin-Madison, Mathematical Research Center.
- Rao, C. R. 1973. Linear Statistical Inference and its Applications (2nd ed.). John Wiley & Sons, New York.
- Stuart, A. 1954. A simple presentation of optimum sampling results. Journal of the Royal Statistical Society, B 16, 239-241.
- Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., Asok, C. 1984. Sampling Theory of Surveys with Applications. Third edition. Ames, Iowa: Iowa State University Press.
- Thompson, M. E. 1997. Theory of Sample Surveys. Chapman & Hall.
- Winston, W. L. 1995. Introduction to Mathematical Programming: Applications and Algorithms. Duxbury Press.